RESIDUAL RANKS TESTING FOR OUTLIERS IN 2^n DESIGNS

Robert Schall
Messerschmitt-Bolkow-Blohm Gmbtt
D-8000 Munchen, Germany
and University of Cape Town

Timothy T. Dunne
University of Cape Town
Republic of South Africa

Technical Report No. 485
June, 1986

ADDITIONAL VARIABLES AND ADJUSTED ESTIMATES WITH ARBITRARY KNOWN VARIANCE-COVARIANCE STRUCTURE

Robert Schall
Messerschmitt-Bolkow-Blohm Gmbtt
D-8000 Munchen, Germany
and University of Cape Town

Timothy T. Dunne*
University of Cape Town
Rondebosch
RSA

Key words and phrases: arbitrary variance; additional variables; extra sum of squares; downdating formulae;

ABSTRACT

When additional variables are fitted in a linear model under arbitrary known variance-covariance structure, the extra sum of squares due to fitting the new variables and adjusted parameter estimates can be computed in an efficient manner without actually explicitly fitting the entire augmented model. When the additional variables are specific dummy variables, downdating formulae are readily obtained, thus generating methods which are well known for the linear model with variance-covariance structure $\sigma^2 I$. Two different methods to downdate a linear model are presented.

* Partial support from University of Minnesota, by sabbatical leave grant from FRD, CSIR, South Africa.

INTRODUCTION

Consider a simple normal sample written as a linear model

$$\begin{bmatrix} Y_1 \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mu + e \qquad (1)$$

The usual test whether a specified observation Y_n (say) is an outlier based on the standardized residual is equivalent to a test of $H_o:\gamma=0$ against $H_A:\gamma\neq 0$ in the adjusted model

$$\begin{bmatrix} Y_1 \\ Y_{n-1} \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \gamma \end{bmatrix} + e \qquad (2)$$

as formulated by John and Draper (1978).

It would now be desirable to also have a nonparametric test for H_o, since in many practical situations normality can not be assumed, even though least squares estimation of μ may be envisaged.

But although (2) can be seen as a straightforward two-sample problem with sample sizes 1 and (n-1) respectively, the application of nonparametric tests based on ranking the data such as Mann-Wilcoxon, Wald-Wolfowitz and Fisher-Yates makes no sense since Y_n is naturally singled out as a possible outlier because it is the largest (or smallest) observation in the first place. Thus ranking the original sample data will not provide any additional information as to whether the observation in question is an outlier.

In the following, however, we present a method to apply the named nonparametric tests for the detection of outliers in a wide class of experimental designs with 2^n observations. The method involves a simple orthogonal transformation of the data and thereafter the routine use of those and Kendall and Stuart, 1973) that the parameter estimates under (1.2) can be obtained without actually explicitly fitting the entire model (1.2). Essentially, the model

$$y = P_{z|x} A\lambda + e; \quad cov(e) = \sigma^2 I \qquad (1.3)$$

is fitted, and the BLUE $P_{z|x}A\hat{\lambda}$ for $P_{z|x}A\lambda$ under (1.3) is computed, which is also the BLUE for $P_{z|x}A\lambda$ under (1.2). Then the extra sum of squares SSA due to fitting A in (1.2) is given by $SSA = (P_{z|x}A\hat{\lambda})'(P_{z|x}A\hat{\lambda})$, and the BLUE $X\hat{\beta}^{(1.2)}$ for $X\beta$ under (1.2) is given by $X\hat{\beta}^{(1.2)} = X\hat{\beta}^{(1.1)} - P_{x|z}A\hat{\lambda}$, i.e. by adjusting the BLUE $X\hat{\beta} = X\hat{\beta}^{(1.1)}$ for $X\beta$ under (1.1).

In the following, we generalize these methods for the case where V is an arbitrary known non-negative definite matrix. Thereafter, we present two different methods for downdating a linear model in this general case, and downdating formulae are readily obtained by taking A in (1.2) to be specific sets of dummy variables.

2. ADJUSTED ESTIMATES AND THE EXTRA SUM OF SQUARES

To clarify the notation we list and number a sequence of linear models, all of them under the common variance-covariance structure $\sigma^2 V$.

(1) $y = X\beta + e$; : the original model

(2) $y = [X:A]\begin{bmatrix}\beta\\ \lambda\end{bmatrix} + e$; : the augmented model, we take $C(A) \subset C([X:V])$

(3) $y = [X:P_{vz|x}A]\begin{bmatrix}\beta\\ \lambda\end{bmatrix} + e$; : a reparametrization of (2)

(4) $y = P_{vz|x}A\lambda + e$; : a reduction of (3).

Lemma; Let $\hat{y}^{(i)}$, $X\hat{\beta}^{(i)}$ and $P_{vz|x}A\hat{\lambda}^{(i)}$ respectively denote the BLU-estimates for the fitted values, $X\beta$ and $P_{vz|x}A\lambda$ in the models in question. Then the following relationships hold, provided that $C(A) \subset C([X:V])$:

as a $LM(Y, X\beta, \sigma^2 I)$ where

$$X = \begin{bmatrix} 1,\ldots,1,0,\ldots,0 \\ 0,\ldots,0,1,\ldots,1 \end{bmatrix}$$

2^{-n} 2^{-n}-times

But the columns of X span the same space as

$$Z = \begin{bmatrix} 1,\ldots,1,\ 1,\ldots,\ 1 \\ 1,\ldots,1,-1,\ldots,-1 \end{bmatrix}$$

The two columns of Z are two columns of T and hence we have $X \in S$.

Example 1.3 By definition a 2^n factorial design where all main effects and interactions are fitted has a design matrix $X \in S$, as well as any smaller model than the full model.

Moreover, a 2^n factorial design in 2 (say) replicates has a design matrix $Z \in S = S^{(n+1)}$, since the 2^n columns of $Z = \begin{bmatrix} X \\ X \end{bmatrix}$ are clearly the first 2^n columns of $T=T^{(n+1)} = (\otimes D)^{n+1}$, the design matrix of the full model of a 2^{n+1} factorial design. Similarly, it can be seen that a 2^n factorial design in 2^r replicates $(r \in x)$ has a design matrix $Z \in S^{(n+r)}$.

Generally, it can be said that if $X_1 \in S^{(n_1)}$ and $X_2 \in S^{(n_2)}$ then $X = X_1 \otimes X_2 \in S^{(n_1+n_2)}$.

In summary, Table 1.4 lists more exhaustively a number of experimental designs where the method outlined in the following section can be applied:

denote respectively the degrees of freedom associated with the hypothesis and the degrees of freedom for error under (1).

Proof:

(a) $P_{vz|x}A$ can be written as $(I - X(X'V^*X)^- X'V^*)A$, and thus (3) is a reparametrization of (2). The equality of the fitted values $\hat{y}^{(2)} = \hat{y}^{(3)}$ follows.

(b) Let Z be a matrix of maximum rank such that $Z'[X:A] = 0$. Then any $y \in C([X:V])$ can be written as

$$
\begin{aligned}
y &= X\xi + A\alpha + VZ\gamma \text{ , for some } \xi,\ \alpha,\ \gamma \\
&= X\xi + (P_{x|vz} + P_{vz|x})\, A\alpha + V\bar{Z}\,\gamma. \qquad (2.1)
\end{aligned}
$$

Clearly,

$$
\begin{aligned}
X\hat{\beta}^{(1)} &= X\lambda + P_{x|vz}A\alpha \\
&= X\hat{\beta}^{(3)}.
\end{aligned}
$$

(c)
$$
\begin{aligned}
\hat{y}^{(3)} &= X\hat{\beta}^{(3)} + P_{x|vz}A\hat{\lambda}^{(3)} \\
&= \hat{y}^{(2)} \qquad \text{, from (a)} \\
&= X\hat{\beta}^{(2)} + A\hat{\lambda}^{(2)} \\
&= X\hat{\beta}^{(2)} + P_{x|vz}A\hat{\lambda}^{(2)} + P_{vz|x}A\hat{\lambda}^{(2)}. \qquad (2.2)
\end{aligned}
$$

Now the first equality in (c) follows from $C(X) \cap C(VZ) = \{0\}$. The second equality is proved using (2.1), provided that model (4) is consistent. In any case, y can be replaced by $\hat{e}^{(1)} = y - X\hat{\beta}^{(1)} = y - X\xi - P_{x|vz}A\alpha$.

(d) using (2.2), (b) and (c) we have

$$
\begin{aligned}
X\hat{\beta}^{(2)} &= X\hat{\beta}^{(3)} - P_{x|vz}A\hat{\lambda} \\
&= X\hat{\beta}^{(1)} - P_{x|vz}A\hat{\lambda}.
\end{aligned}
$$

$$Y = \begin{bmatrix} X_{(-)} & : & 0 \\ X_m & : & 1 \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + e \tag{3}$$

and we test H_o: $\gamma = 0$.

Since $2^{-n/2}T$ is an orthogonal matrix, we can transform the LM(3) by $2^{-n/2}T$ to obtain the LM

$$2^{-n/2}TY = \tilde{Y} = [\tilde{X}:c] \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \tilde{e} \tag{4}$$

where c denotes the last column of $2^{-n/2}T$.

The column space of X is spanned by some q columns of T and thus X is orthogonal to (m-q) columns of T, or equivalently to (m-q) rows of T since T is symmetric.

Possibly after some rearrangement of the transformed data $\tilde{Y}_1, \ldots \tilde{Y}_m$, the LM(4) can be written as

$$\tilde{Y} = \begin{bmatrix} A_q & : & \\ & & c \\ 0 & : & \end{bmatrix} \begin{bmatrix} \beta \\ \\ \gamma \end{bmatrix} + \tilde{e} \tag{5}$$

where A_q is a matrix of order (q×p) of full row rank.

Now we drop the first t transformed observations $\tilde{Y}_1, \ldots, \tilde{Y}_q$ from the model to obtain

$$\begin{bmatrix} \tilde{Y}_{q+1} \\ \tilde{Y}_m \end{bmatrix} = \begin{bmatrix} c_{t+1} \\ c_m \end{bmatrix} \gamma + \tilde{e}_{(-t)} \tag{6}$$

But the i-th component c_i of c equals either $+2^{-n/2}$ or $-2^{-n/2}$ $(i=1, \ldots, 2^n)$, and thus we can see (6) as a two-sample problem.

By construction of T, the transformed observations $\tilde{Y}_2, \ldots \tilde{Y}_m$ in (4) are uncorrelated and identically distributed. If only a general mean is fitted in X, $\tilde{Y}_1$ is dropped while proceeding from (5) to (6), and the observations

$$\hat{e}^{(2)} = y - X\hat{\beta}^{(2)} - A\hat{\lambda}^{(2)}$$

$$A\hat{\lambda}^{(2)} = \hat{e}^{(2)} - y + X\hat{\beta}^{(2)} \qquad (2.3)$$

which implies that $A\lambda\hat{\ }(2) \in C([X:V])$, w.p.1. Thus the results (d) through (h) of the lemma hold for any A when data and model are consistent. If $C(A) \subset C([X:V])$ is not satisfied, a possible course of action is to reparametrize the $A\lambda$-part of model (2) so that

$$A\lambda = A^*\lambda^* = [A_1^* : A_2^*] \begin{bmatrix} \lambda_1^* \\ \lambda_2^* \end{bmatrix}, \text{ for all } \lambda, \lambda^*$$

where $C(A_1^*) \subset C([X:V])$ and $C(A_2^*) \cap C([X:V]) = \{0\}$. Then, using (2.3) we conclude that the BLUE $A_1^*\hat{\lambda}_2^*$ for $A_2^*\lambda_2$ is zero w.p.1. Thus it is sufficient to fit the variables A_1^* only, for which the lemma can be applied.

3. DOWNDATING OF A LINEAR MODEL

Suppose the variables A in the model (1.2) are respectively taken to be the dummy variables

$$A = \begin{bmatrix} 0 \\ I_k \end{bmatrix}, \text{ or} \qquad (3.1)$$

$$A = V_2 = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \begin{bmatrix} 0 \\ I_k \end{bmatrix}, \text{ where } V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \qquad (3.2)$$

Then fitting those dummy variables is equivalent to the reduction of the LM(1.1) to the models

$$y_1 = X_1\beta + e_1 \ ; \qquad \text{cov}(e_1) = \sigma^2 V_{11} \qquad (3.3)$$

2^{ℓ}.

For example, in the case of two possible outliers, possibly after some rearrangment of the data, (7) can be written as

$$\begin{bmatrix} \tilde{Y}_{q+1} \\ \\ \\ \\ Y_m \end{bmatrix} = 2^{-n/2} \begin{bmatrix} 1 & 1 \\ & \\ 1 & -1 \\ -1 & 1 \\ \vdots & \vdots \\ -1 & -1 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix} + e$$

which may be treated as a 4-sample problem. Note that if we know that the estimated residual $\hat{e}_m$ is larger than $\hat{e}_{m-1}$ (say), we would test

$$H_o: \gamma + \delta = \gamma - \delta = -\gamma + \delta = -\gamma - \delta$$

against

$$H_A: \gamma + \delta > \gamma - \delta > -\gamma + \delta > -\gamma - \delta$$

and concerning this type of alternative, Kendall and Stuart (1973, p. 524) conjecture that the U-test based on the number of inversions between the samples might be more efficient than the H-test based on the rank sums of the samples.

4. THE CONVERSE PROBLEM: SMALL SAMPLE SIZES IN SLIPPAGE PROBLEMS

Consider for example a two-sample problem where the size of the one sample is very small, just one or two observations (say). This situation may arise quite frequently when it is difficult to obtain experimental material for the application of a certain treatment, whereas a control group (the second sample) of reasonable size might be easily available. In such a case nonparametric tests are known to have relatively low efficiency, and if the total sample size is small a nonparametric test might not even obtain size α

for a given α. But with the method of the foregoing sections the suggested test can be applied if the total sample size is $m=2^n$.

If the first sample consists only of one observation we may treat this observation as an outlier and test for significance as outlined in Section 2.

In the case of sample size 2 for the first sample we have the model

$$\begin{bmatrix} Y_1 \\ \\ Y_{m-2} \\ Y_{m-1} \\ Y_m \end{bmatrix} = \begin{bmatrix} 1 & : & 0 \\ \\ 1 & : & 0 \\ 1 & : & 1 \\ 1 & : & 1 \end{bmatrix} \begin{bmatrix} \mu_2 \\ \mu_1 \end{bmatrix} + e$$

After transformation by $2^{-n/2}T$ and possibly some rearrangement of the transformed data we arrive at

$$\begin{bmatrix} \tilde{Y}_1 \\ : \\ : \\ : \\ : \\ : \\ : \\ \\ \tilde{Y}_m \end{bmatrix} = 2^{-n/2} \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \\ 0 \\ -2 \\ \\ -2 \end{bmatrix} \mu_2 + \tilde{e}$$

We can now apply a 3-sample rank test (preferably based on the H-statistic).

5. ON THE EFFICIENCY OF THE OUTLIER-TESTS

The non-null distribution of the test-statistics based on the ranks of the transformed observations $\tilde{Y}_{q+1},\ldots\tilde{Y}_m$ is certainly intractable in the general case, but the efficiency of the tests can be judged by a comparison with the normal alternative. If the observations Y in the original model (3) follow a normal distribution, then so do the transformed observations $\tilde{Y}$ in models (4) through (6). Further, the observations $\tilde{Y}$ are independently distributed, and $\tilde{Y}_{q+1},\ldots,\tilde{Y}_m$ are in addition identically distributed under H_o. Thus the results on the asymptotic relative efficiency (ARE) of the various nonparametric tests, known from the literature (see e.g. Kendall and Stuart, 1973), can be applied straightforwardly, noting the invariance of the normal theory F- or t-test under a transformation of a linear model.

We may note that, as in the case of 1 outlier as in model (6), the nonparametric test is based on m-q observations, whereas the corresponding t-test has degrees of freedom m-q-1. Thus while one observation is gained by performing the nonparametric test in question, it is the partitioning of the ranked residuals by the vector c of (4) which results in a corresponding loss of efficiency. In fact, with A_q as given in (5) and q=1 we observe that it is the information associated with the first observation $\tilde{Y}_1$ which is lost, since $\tilde{Y}_1$ appears in all "contrasts" $\tilde{Y}_1,\ldots,\tilde{Y}_m$ with a + sign. The observation $\tilde{Y}_1$ could take an arbitrary value without altering the resulting nonparametric test-statistic. Essentially for q>1 we perform the test using the observations $\tilde{Y}_2,\ldots\tilde{Y}_m$ only, and thus m-q residuals only are required for application of the nonparametric tests. The zero residuals in (5) may be allocated on a tied-ranks basis to the corresponding subsets determined by their entries in the vector c.

REFERENCES

Cook, R.D. and Weisberg, S. (1982). Residuals and Influence in Regression. Chapman and Hall, New York.

John, J.A. and Draper, N.R. (1978). On testing for two outliers or one outlier in two-way tables. Technometrics, 20, 69-78.

Kendall, M.G. and Stuart, A. (1973). The Advanced Theory of Statistics, Vol. 2, 3rd Edition. Griffin, London.

Searle, S.R. (1971). Linear Models. Wiley, New York.

Rao, C.R. (1971). Unified theory of linear estimation. Sankya, Ser. A, 33, 371-394.

Rao, C.R. (1973). Linear Statistical Inference and its Applications. Wiley, New York.

Rao, C.R. (1976). Estimation of parameters in a linear model. Annals of Statistics, 4, 1023-1037.

Schall, R. and Dunne, T.T. (1985). On outliers, recursive residuals, and missing data estimation for arbitrary known variance-covariance structure. Technical report No. 223, Department of Statistics, Stanford University, Stanford, California.

Schall, R. and Dunne, T.T. (1986). Augmenting a linear model under arbitrary covariance structure. Technical report No. 226, Department of Statistics, Stanford University, Stanford, California.

Zyskind, G. & Martin, F.B. (1967). A general Gauss-Markoff theorem for linear models with arbitrary non-negative covariance structure. SIAM Journal of Applied Mathematics, 17, 1190-1202.